Predicting Customer Churn in a Telecommunications Company

Objective

The primary aim of this project is to develop a predictive model that identifies customers at risk of churning, enabling proactive measures to retain them and reduce churn rates.

Understanding Data

- Customers who left within the last month – the column is called Churn

- Services that each customer has signed up for – phone, multiple lines, internet, online security, online backup, device protection, tech support, and streaming TV and movies

- Customer account information – how long they've been a customer, contract, payment method, paperless billing, monthly charges, and total charges

- Demographic info about customers – gender, age range, and if they have partners and dependents

EXPLORATORY DATA ANALYSIS (EDA)
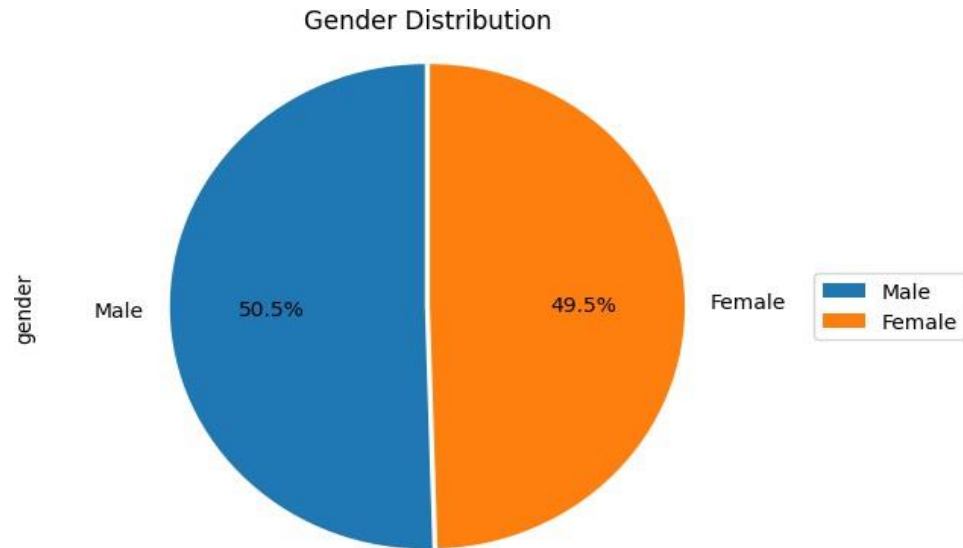
Churn Distribution

The observed churn rate stands at 26.5%, indicating that a notable portion of customers discontinued services. In contrast, 73.5% retained services, which is a significant majority. Understanding the reasons behind this churn is crucial for devising effective retention strategies. Factors like service satisfaction, pricing, or contract terms could contribute to this churn rate.
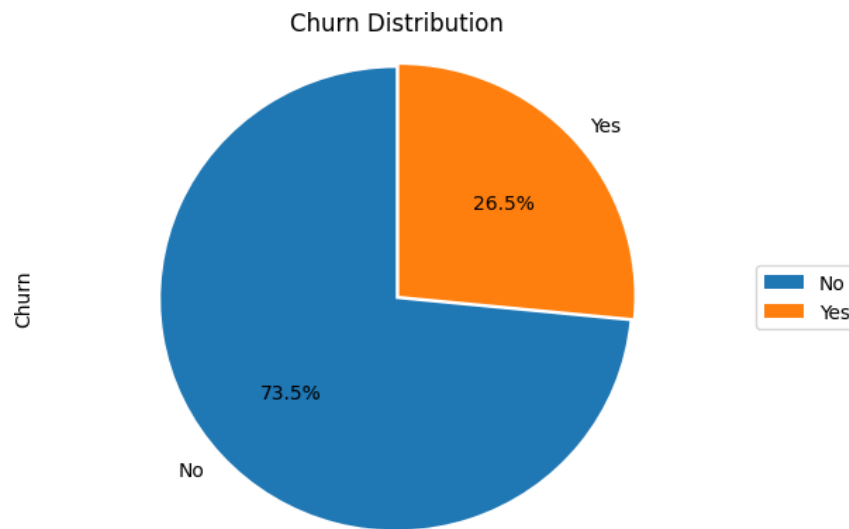

Demographic Insights

Gender Distribution

The dataset showcases an almost equal distribution between male and female customers. This balance in gender representation suggests that the telecommunications services cater to a diverse customer base.
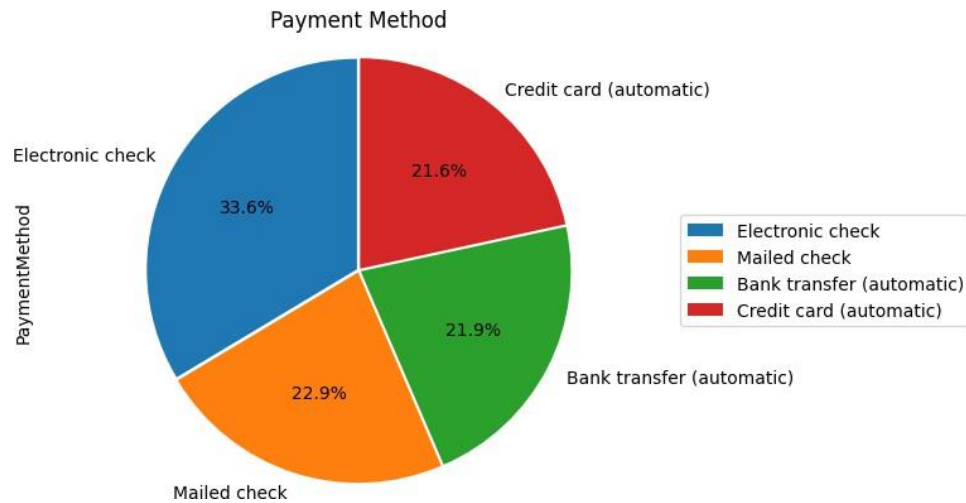
Gender Distribution

Age Group Distribution

With 1142 identified as senior citizens and 5901 as young adults, the dataset reflects a mix of different age demographics. Understanding the needs and preferences of these diverse age groups could aid in customizing services or promotions tailored to specific age brackets, potentially impacting churn rates.
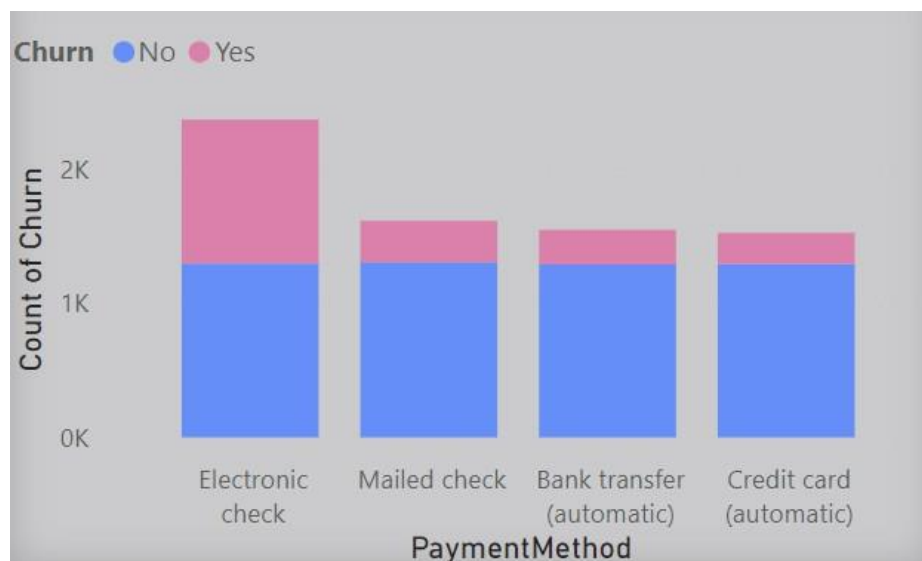

Senior Citizen Counts
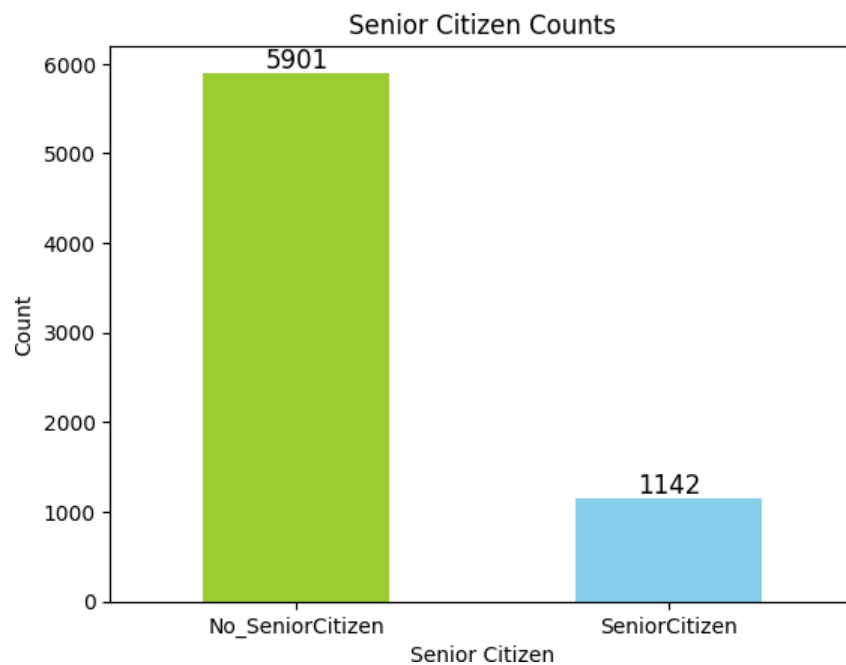
Billing and Payment Preference

Different distributions observed among electronic check, Mailed Check, Bank transfer(automatic) and credit card (automatic).
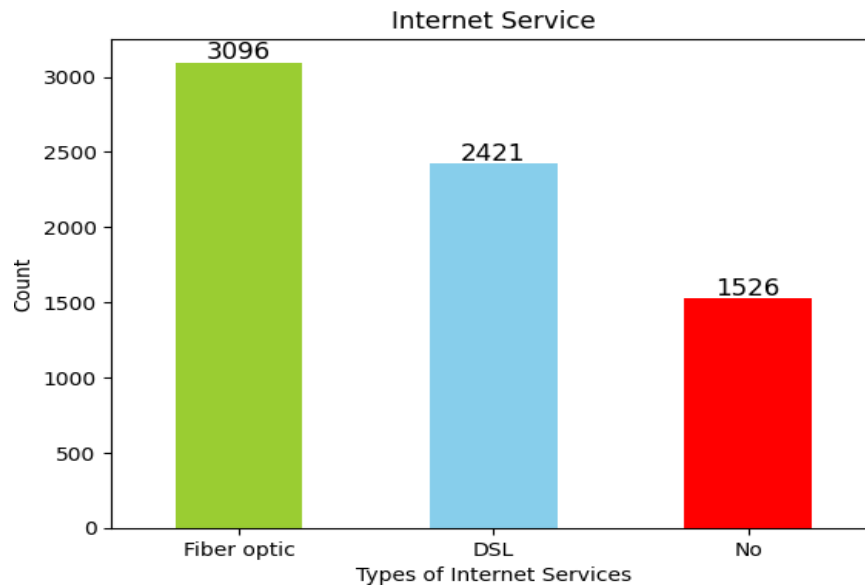

Insights

We found chunk rate is very high for electronic check payment method in comparison of other payment methods.

Service-Related Insights

Internet Service Types

Different distributions observed among Fiber Optic, DSL, and customers without internet service.
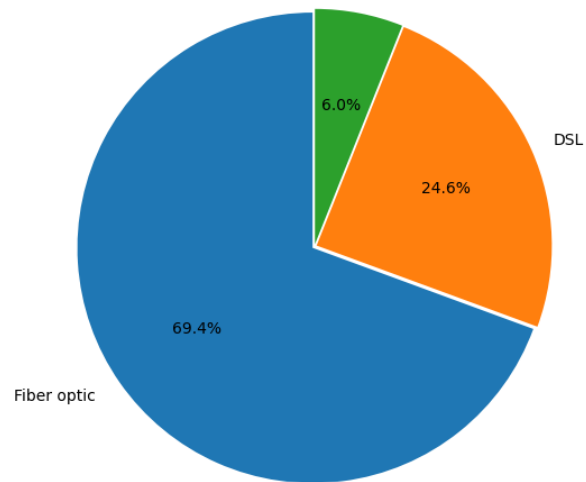

Insights

- Fiber optic service is preferred by a substantial number of customers (3096).
- Churn rates among different internet service types:
- Fiber optic: 1297 out of 3096 churned.
- DSL: 459 out of 2421 churned.
- No internet service: 113 out of 1526 churned.

A significant number of customers opt for Fiber Optic internet services. This preference might indicate a higher demand for faster internet speeds or specific service features associated with fiber Optic plans. Analyzing churn rates among different internet service types could unveil service-related factors influencing customer retention.

Churn Distribution for Fiber Optic Customers

Service Bundles

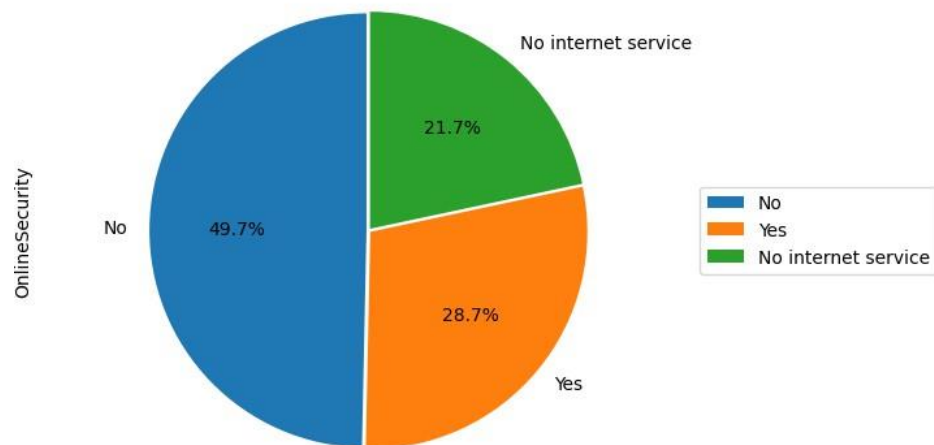## Online Backup

Distribution for online Backup

Distribution For Online Security
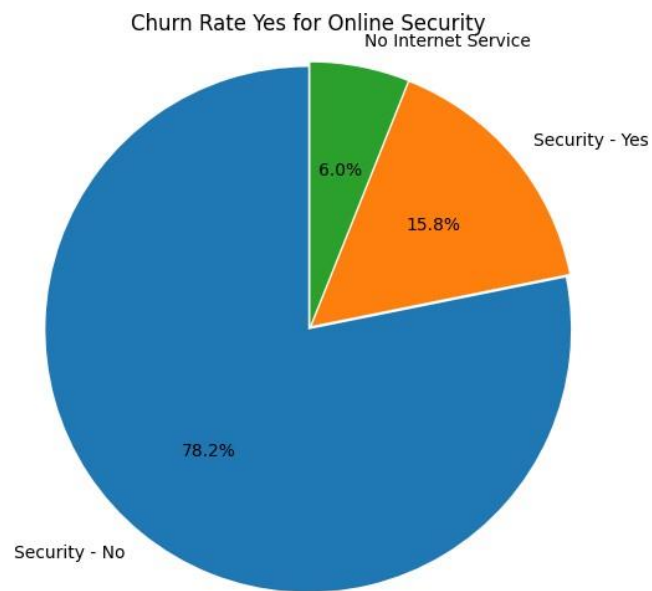
Insights

Churn Rate Yes for Online Security

- The number of people who left the service with no online security are 1461 out of 3498
- The number of people who left the service with online security are 295 out of 2019
- The number of people who left the service with no internet service are 113 out of 1526

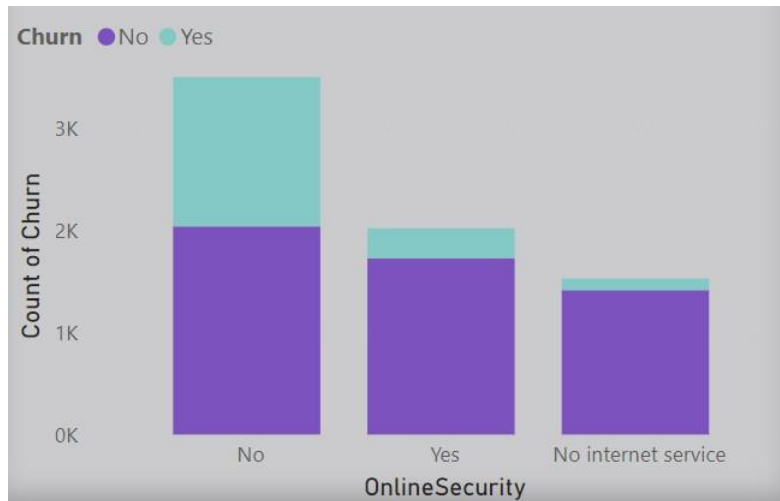**Tech Support**
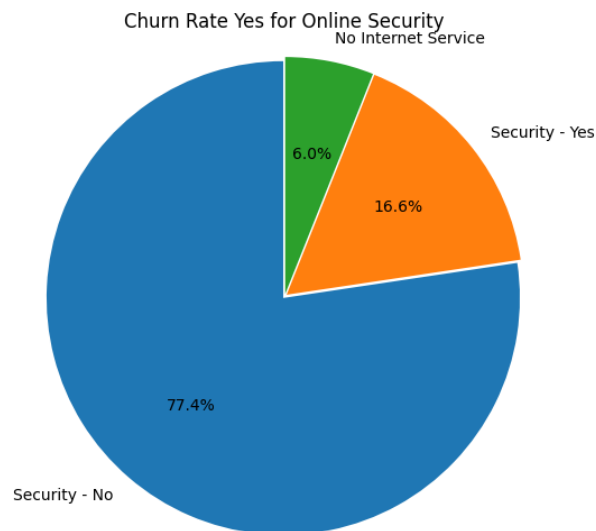
Distribution for Online Security –

Insights

The number of people who left the service with no online security are 1461 out of 3498.

The number of people who left the service with online security are 295 out of 2019

The number of people who left the service with no internet service are 113 out of 1526.

Contract Types

Different types of Contract length are –
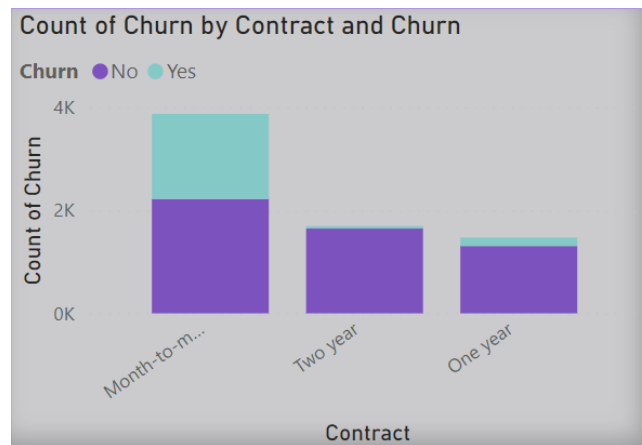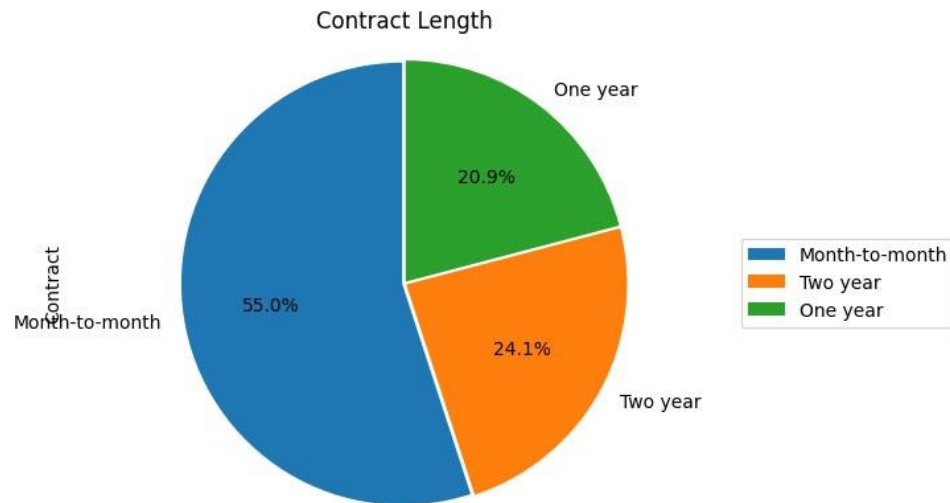
Insight Churn Rate for Online Security

Churn Rate for different Contract Length
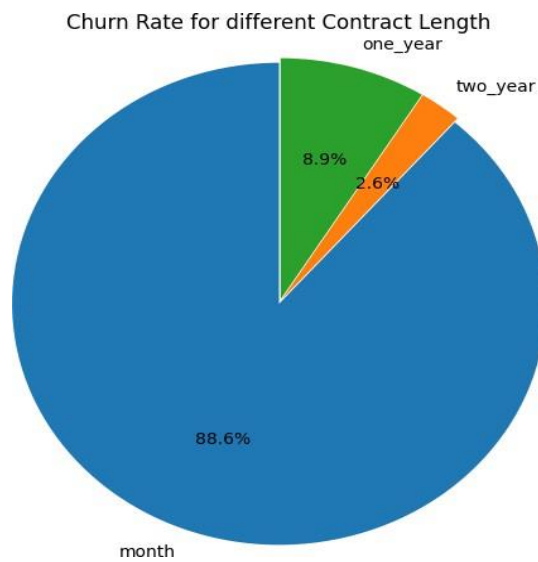
- The number of people who left the service with month-to-month contact are 1655 out of 3875
- The number of people who left the service with two-year contract are 48 out of 1695
- The number of people who left the service with one year contract are 166 out of 147

Tenure Analysis

Our average Tenure length is 29 days.

Churn Rate for more than average tenure length
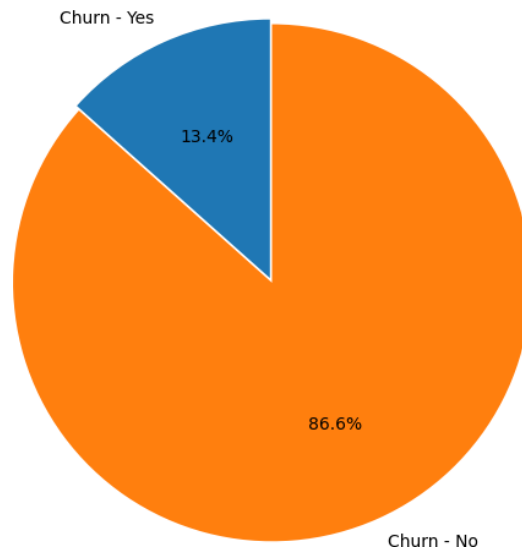
At Average tenure length of 29 Days churn rate is very low.
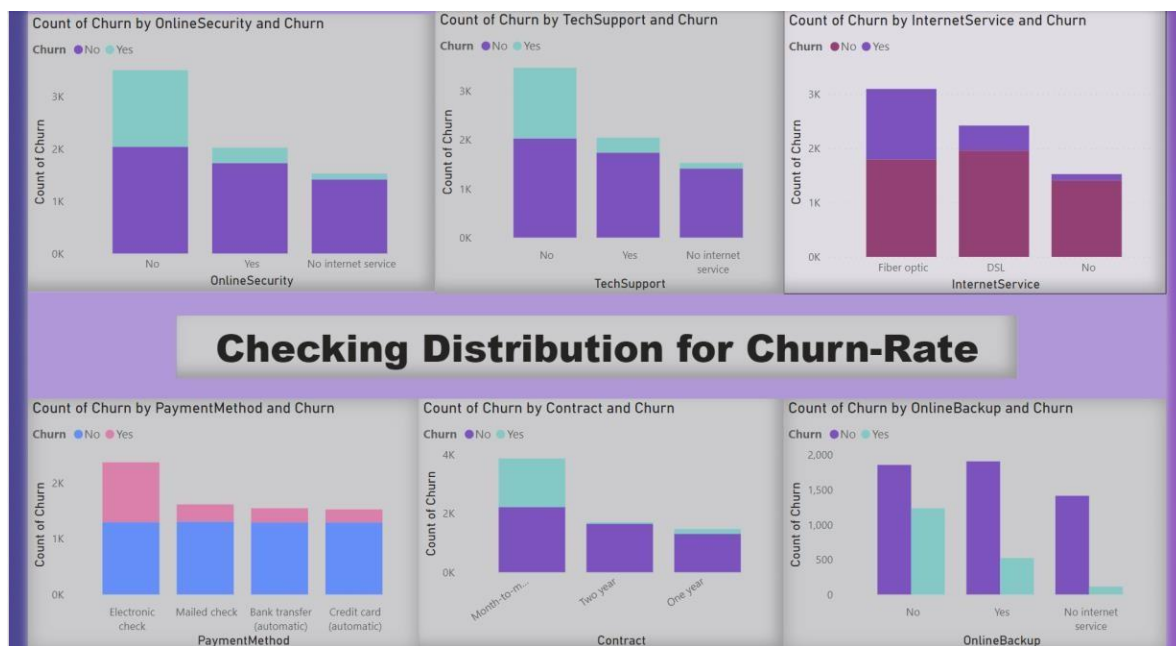
Correlation Matrix Insight

A heatmap visualizes relationships between features and churn. Notable negative correlations exist with tenure and certain service subscriptions, indicating their potential to reduce churn rates. Conversely, a moderate positive correlation with Fiber Optic internet service implies a higher likelihood of churn.
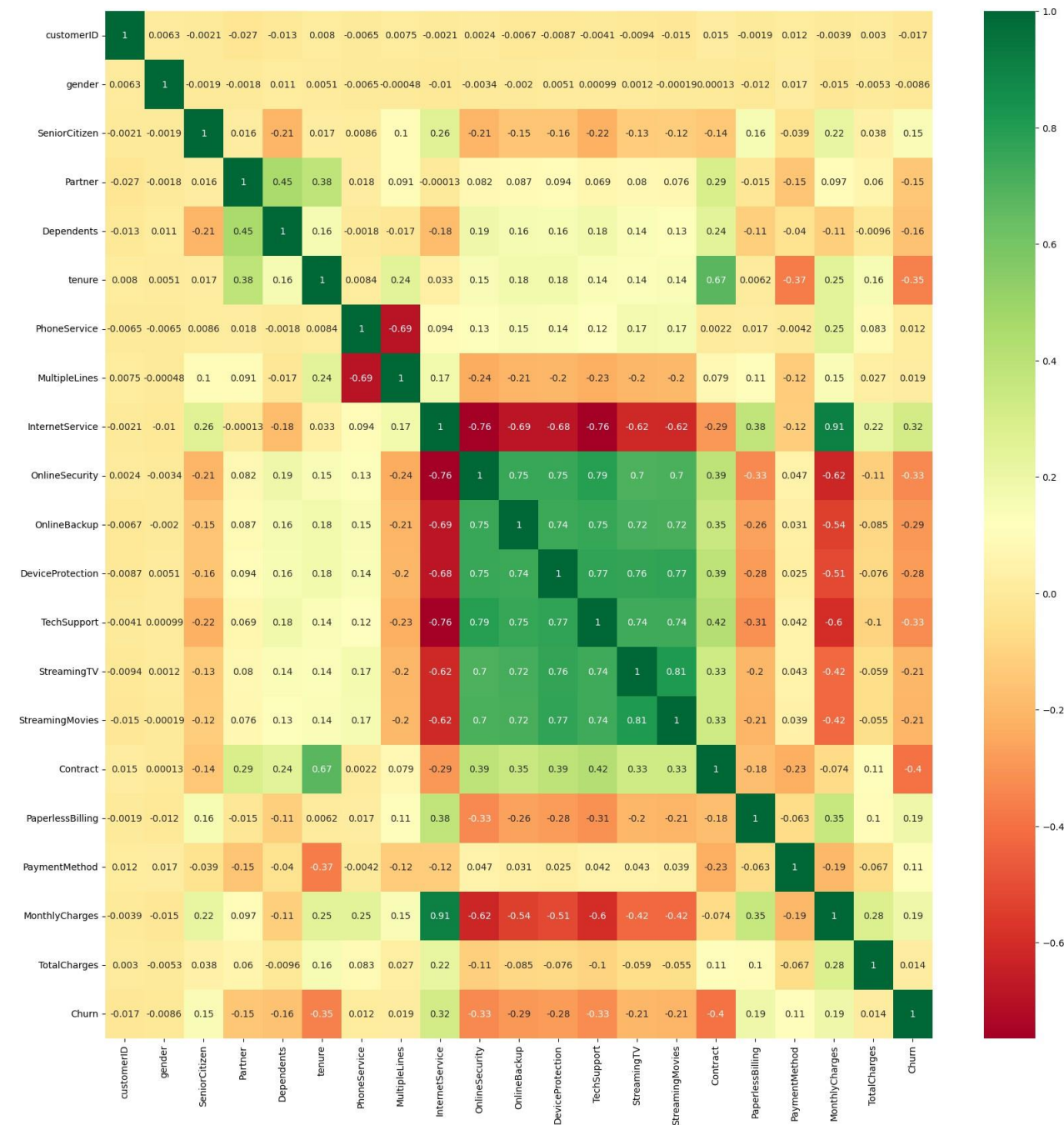
- 1: Perfect positive correlation. When one variable increases, the other variable also increases proportionally.
- 0: No correlation. There's no linear relationship between the variables.
- -1: Perfect negative correlation. When one variable increases, the other variable decreases proportionally.
- Darker shades indicate stronger correlations (either positive or negative).
- Closer to 1 or -1 indicates a stronger relationship, whereas values closer to 0 signify weaker or no relationship.

This suggests an inverse relationship between those variables. For instance, if one variable represents customer satisfaction and another represents churn, a negative correlation would suggest that higher satisfaction is associated with lower churn rates.

Customer ID (-0.017447):

Analysis reveals no substantial correlation between Customer ID and churn, indicating no discernible impact on customer attrition.

Gender (-0.008612)

A minimal correlation exists between gender and churn, suggesting a negligible influence on customer attrition.

Senior Citizen (0.150889)

A slight positive correlation is observed, indicating that older customers demonstrate a marginally higher likelihood of churning compared to younger counterparts.

Partner (-0.150448), Dependents (-0.164221):

Moderate negative correlations are evident, suggesting that customers with a partner or dependents exhibit a tendency to remain with the service for a longer duration, reducing the likelihood of churn.

Tenure (-0.352229)

A strong negative correlation is observed, indicating that customers with longer tenure demonstrate significantly lower churn rates, emphasizing the importance of tenure in reducing attrition.

Internet Service (0.316846):

A notable positive correlation is identified, suggesting that customers using Fiber Optic service exhibit a significantly higher likelihood of churning compared to other service types.

Contract (-0.396713):

A strong negative correlation is evident, indicating that customers with longer-term contracts experience notably reduced churn rates, emphasizing the effectiveness of long- term contracts in retaining customers.

Online Security, Online Backup, Device Protection, Tech Support, Streaming-TV, Streaming-Movies

Negative correlations across these services suggest that customers availing these additional services tend to display lower churn rates, indicating the potential impact of these services in enhancing customer retention.

Monthly Charges, Paperless Billing

Slight positive correlations are observed, indicating that higher monthly charges and the adoption of paperless billing contribute marginally to increased churn rates among customers.

These insights provide valuable guidance for devising strategies aimed at mitigating churn by focusing on customer tenure, service enhancements, contract structures, and billing preferences.

CONCLUSION

Fiber Optics Users and Churn

Users of Fiber Optic services display a higher propensity towards churning.

Tenure Length and Churn

A minimum tenure length of 29 days demonstrates significantly lower churn rates, highlighting the importance of encouraging longer tenure among customers.

Online Security Impact on Churn

Among users who left the service:

- Those without online security services: 1461 out of 3498.
- Those with online security services: 295 out of 2019.
- Those without internet service: 113 out of 1526.

Customers without security services exhibit notably higher churn rates compared to those with security services.

Contract Types and Churn

Among customers who left the service:

- Month-to-month contract users: 1655 out of 3875.
- One-year contract users: 166 out of 1473.
- Two-year contract users: 48 out of 1695.

Month-to-month contract users show a higher churn rate compared to longer-term contract users.

Tech Support Influence on Churn

Among customers who left the service:

- Those without tech support: 1446 out of 3473.
- Those with tech support: 310 out of 2044.
- Those without internet service: 113 out of 1526.

Customers lacking tech support services demonstrate a higher churn rate compared to those with tech support.

These observations underline the impact of different service usage patterns, contract lengths, and support services on customer churn rates, emphasizing the need to address these factors for better retention strategies.

RECOMMENDATION

Fiber Optics Users and Churn

Recommendation: Focus on understanding the pain points of Fiber Optic users and develop retention strategies, such as providing better service quality or personalized offers, to reduce churn among this group.

Tenure Length and Churn

Recommendation: Encourage and incentivize customers to stay longer by offering loyalty rewards, exclusive benefits, or discounts for extended tenure periods, aiming to minimize churn.

Online Security Impact on Churn

Recommendation: Emphasize the importance of online security features to customers and potentially offer introductory packages or incentives to encourage adoption, aiming to reduce churn rates associated with a lack of security services.

Contract Types and Churn

Recommendation: Introduce promotional offers or perks for longer-term contracts to entice customers to opt for extended contract durations, thereby lowering the churn rate associated with month-to-month contracts.

Tech Support Influence on Churn

Recommendation: Enhance the visibility and value proposition of tech support services, highlighting their importance in ensuring a smooth customer experience, thereby potentially reducing churn rates linked to a lack of tech support.

Implementing these recommendations can help mitigate churn rates by addressing specific pain points related to service usage, tenure, security, contract lengths, and support services, ultimately leading to improved customer retention and satisfaction.

BUILDING THE CHURN PREDICTION MODEL

Data Collection and Preprocessing

Data Collection

The Telco Customer Churn dataset from Kaggle was utilized for this project. The dataset comprises information about telecommunications customers, including demographics, services subscribed, contract details, and churn status.

Preprocessing Steps

Handling Missing Values

Missing values within the dataset were identified and addressed using appropriate strategies:

Detection: Recognized the presence of missing values within various features.

Correction: Although the current dataset doesn't contain missing values, if they were present, we would employ imputation methods (like mean, median, or mode) to handle missing values in numerical features.

Exclusion: Contemplated removing records or features exhibiting a substantial proportion of missing values.

Encoding Categorical Variables

Categorical variables were encoded to transform them into a numerical format suitable for modeling:

Binary conversion: Applied one-hot encoding to categorical variables with multiple categories to create binary columns for each category.

Label Encoding: Utilized label encoding for ordinal categorical variables, preserving the ordinal relationships between categories.

Feature Scaling

For some machine learning algorithms, feature scaling was applied to standardize or normalize numerical features:

Standard Scaling: Scaled numerical features to have a mean of zero and a standard deviation of one.

Normalization: Normalized features to a specific range to prevent features with larger scales from dominating during model training.

Data Preparation for Analysis

After preprocessing, the dataset was suitably transformed and prepared for analysis:

**Train-Test Split**: Segregated the dataset into training and testing sets to train and evaluate the models.

Feature Selection: Identified and selected relevant features based on their impact on predicting customer churn.

Machine Learning Algorithms

- Considered and implemented models such as logistic regression, random forests, gradient boosting, Decision Tree Classifier, SVC, GaussiannNB etc.

- Trained models using the preprocessed dataset.

- Fine-tuned hyperparameters using techniques like cross-validation or grid search

Model Training Result

- 'Logistic Regression': 0.7570048309178744,

- 'Decision Tree': 0.8893719806763285,

- 'Random Forest': 0.9135265700483092,

- 'Gradient Boosting': 0.8014492753623188,

- 'SVM': 0.5743961352657004,

- 'K-Nearest Neighbors': 0.6714975845410628,

- 'Naive Bayes': 0.7333333333333333
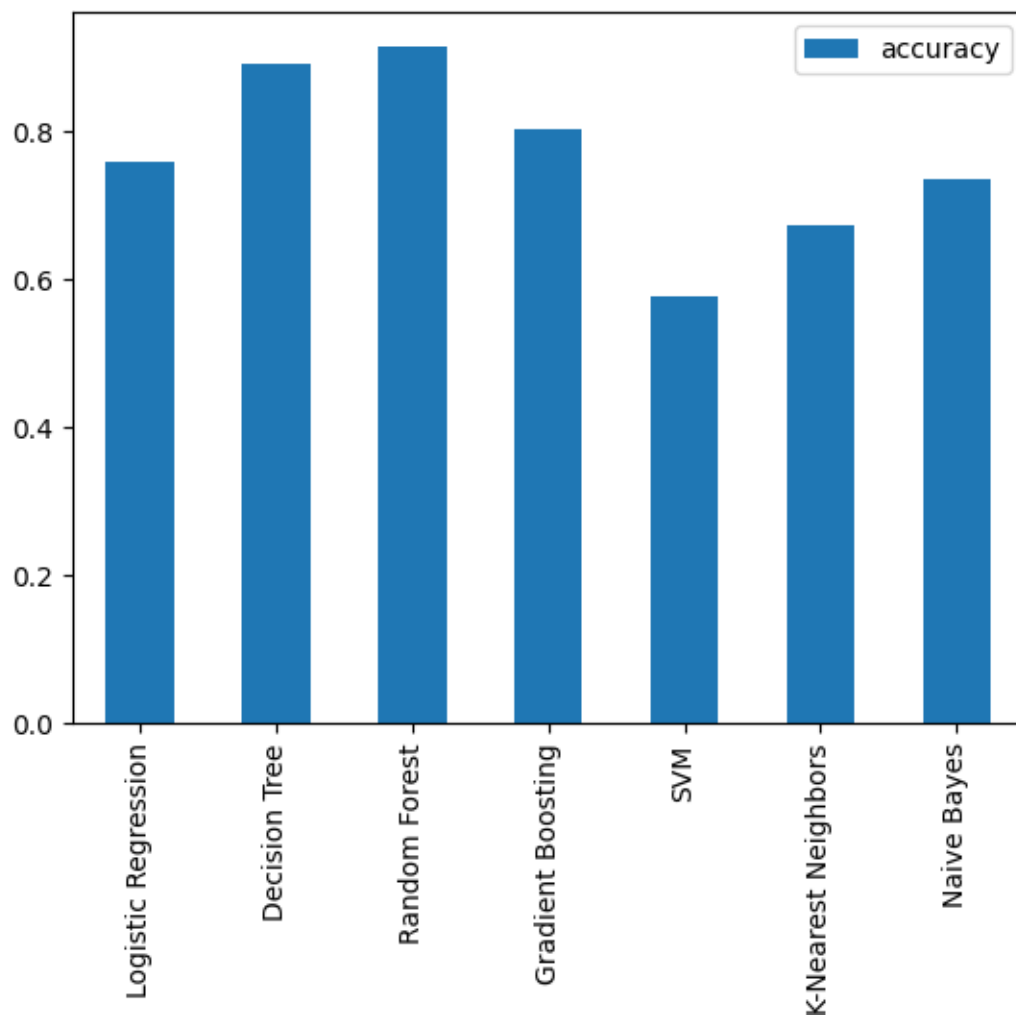
MODEL EVALUATION

Performance Metrics

Evaluated model performance using appropriate metrics: Accuracy, precision, recall, F1-score, ROC-AUC, etc.

Random Forest classifier gives the best accuracy and result and good performance compared the performance of different models.

```
▼ RandomForestClassifier
RandomForestClassifier()
```

Evaluation

- Accuracy: 0.9159
- Precision: 0.8743
- Recall: 0.9698
- F1-score: 0.9196